WHAT IS CLAIMED IS:

1. An information retrieval system for retrieving information a user seeks from a plurality of documents, comprising:

5   document storage means for storing the plurality of documents;

feature amount extraction means for extracting a feature amount of each of the plurality of documents stored in the document storing means;

10   clustering means for classifying the plurality of documents stored in the document storage means into a plurality of clusters based on the extracted feature amounts so that each cluster includes one document or a plurality of documents having feature amounts similar to each other;

15   document retrieval means for retrieving a document satisfying a retrieval condition input by the user among the plurality of documents stored in the document storage means; and

interface means for presenting the retrieved document
20 together with the rest of documents included in a cluster to which the retrieved document belongs if the cluster includes a plurality of documents, as retrieval results.

2. The information retrieval system of Claim 1, wherein, from each of the plurality of documents stored in the docu
25 ment storage means, the feature amount extraction means ex-

31

tracts, as the feature amount, a feature vector including as elements a pair of a term appearing in the document and a weight with which the term characterizes the document.

3. The information retrieval system of Claim 1, wherein the clustering means adopts a way of clustering that provides the largest number of clusters each including a plurality of documents.

4. The information retrieval system of Claim 1, further comprising cluster label preparation means for preparing a plurality of cluster labels respectively representing the contents of the plurality of clusters,

wherein the interface means presents a cluster label representing the contents of the cluster to which the re-trieved document belongs among the plurality of cluster la-bels prepared, together with the retrieval results.

5. The information retrieval system of Claim 4, wherein, for each of the plurality of clusters, the cluster label preparation means selects one or a plurality of terms charac-terizing the cluster from all documents belonging to the cluster as the cluster label.

6. The information retrieval system of Claim 4, wherein, for each of the plurality of clusters, the cluster label preparation means selects one sentence characterizing the cluster from all documents belonging to the cluster as the cluster label.

7. The information retrieval system of Claim 4, further comprising document label preparation means for preparing a plurality of document labels respectively representing the contents of the plurality of documents stored in the documents storage means,

wherein the interface means presents document labels representing the contents of the documents included in the cluster to which the retrieved document belongs among the plurality of document labels prepared, together with the retrieval results.

8. The information retrieval system of Claim 7, wherein, for each of the plurality of documents stored in the document storage means, the document label preparation means selects one sentence characterizing the document from all sentences in the document as the document label.

9. The information retrieval system of Claim 1, wherein the plurality of documents includes a plurality of question documents and a plurality of answer documents associated with each other,

the retrieval condition is a natural language user question,

the feature amount extraction means extracts a feature amount of each of the plurality of answer documents stored in the document storage means to enable the plurality of answer documents to be classified into a plurality of clusters by

the clustering means,

the information retrieval system further comprises similarity operation means for calculating similarity between each of the plurality of question documents stored in the document storage means and a document of the user question,

the document retrieval means retrieves a question document having high similarity among the plurality of question documents stored in the document storage means based on the calculated similarity, and retrieves an answer document associated with the retrieved question document among the plurality of answer documents stored in the document storage means, and

the interface means presents the retrieved answer document together with the rest of answer documents included in a cluster to which the retrieved answer document belongs if the cluster is composed of a plurality of answer documents, as the retrieval results.

10. The information retrieval system of Claim 9, wherein the interface means presents the retrieval results to the user.

11. The information retrieval system of Claim 10, wherein the interface means receives selection of an answer document among the presented retrieval results by the user, and

the information retrieval system further comprises docu-

ment updating means for retrieving a question document asso-
ciated with the selected answer document among the plurality
of question documents stored in the document storage means,
and, if the similarity between the retrieved question docu-
5   ment and the document of the user question is lower than a
predetermined value, newly storing the document of the user
question in the document storage means in association with
the selected answer document.

12. The information retrieval system of Claim 9, wherein
10   the interface means presents the retrieval results to an ex-
pert together with the document of the user question, and
presents an answer document selected by the expert among the
presented retrieval results to the user.

13. The information retrieval system of Claim 12, fur-
15   ther comprising document updating means for retrieving a
question document associated with the selected answer docu-
ment among the plurality of question documents stored in the
document storage means, and, if the similarity between the
retrieved question document and the document of the user
20   question is lower than a predetermined value, newly storing
the document of the user question in the document storage
means in association with the selected answer document.

14. The information retrieval system of Claim 9, wherein
the interface means presents the retrieval results to an ex-
25   pert together with the document of the user question, and

presents to the user a natural language answer document input
by the expert with reference to the presented retrieval re-
sults.

15. The information retrieval system of Claim 14, fur-
ther comprising document updating means for newly storing the
document of the user question and the input answer document
in association with each other in the document storage means
if the similarity between each of the plurality of answer
documents stored in the document storage means and the input
answer document is lower than a predetermined value.

16. An information retrieval system for retrieving in-
formation a user seeks from a plurality of documents, com-
prising:

document storage means for storing a plurality of ques-
tion documents and a plurality of answer documents associated
with each other;

similarity operation means for calculating similarity
between each of the plurality of question documents stored in
the document storage means and a document of a user question
when the user question in natural language is input by the
user;

document retrieval means for retrieving a plurality of
question documents having high similarity among the plurality
of question documents stored in the document storage means
based on the calculated similarity, and retrieving answer

documents associated with the respective retrieved question documents among the plurality of answer documents stored in the document storage means; and

interface means for presenting to an expert the plurality of retrieved answer documents together with the document of the user question, and presenting to the user an answer document selected from the presented retrieval results by the expert or a natural language answer document input by the expert with reference to the presented retrieval results.

17. The information retrieval system of Claim 16, further comprising document updating means for retrieving a question document associated with the selected answer document among the plurality of question documents stored in the document storage means, and if the similarity between the retrieved question document and the document of the user question is lower than a predetermined value, newly storing the document of the user question in the document storage means in association with the selected answer document.

18. The information retrieval system of Claim 16, further comprising document updating means for newly storing the document of the user question and the input answer document in association with each other in the document storage means if the similarity between each of the plurality of answer documents stored in the document storage means and the input answer document is lower than a predetermined value.

ADD A1

37